

Obligatorisk læringsaktivitet i Flow 5, Dataanalyse



Afleveringsfrist:

torsdag den 24. marts 2022, kl. 16

Udarbejdet af:

Wulf & Baum

Indhold

Opgave 1 – MongoDB, Events og opsplittning af data.....	3
Opgave 1.1 – Opsætning af database.....	3
Opgave 1.2 – Indlæs data til R.....	3
Opgave 1.3– Kend din maskine	3
Opgave 2 – Beskrivende statistik og visualisering	4
Opgave 2.1 – Afleveringer.....	4
Opgave 2.2 – Skud.....	4
Opgave 3 – Kvinder, mænd og etik	5
Opgave 3.1 – Forskel i events mellem kvinder og mænd	5
Opgave 3.2 – Investering i kvindefodbold.....	6
Opgave 3.3 – Etiske overvejelser.....	6

Opgave 1 – MongoDB, Events og opsplitting af data

I har allerede indlæst kvinde-kampene fra Statsbomb. Og har noget R-kode, der kan læse data fra mongo.

Opgave 1.1 – Opsætning af database

I skal nu tilføje en attribut til de kampe I har fået indlæst i Mongo så I kan skelne dem fra de herrekampe som I efterfølgende skal indlæse – hint: updateMany

I skal dernæst læse alle herrekampe ind i mongo vha et script som I finder på github.com/cphstud/DALF22W11.

Når kampene er indlæst i mongo sætter I attributterne {gender.name:"male", gender.id:1} på herrekampene – hint: \$exists:false

Opgave 1.2 – Indlæs data til R

Modificer den R-kode som I har brugt til at læse og filtrere smk-data med således at I kan indlæse alle kvindekamp og tilsvarende antal herrekampe.

Opgave 1.3– Kend din maskine

Lav et skærmdump hvor du viser hvor du vil stoppe og starte mongo. Vis også hvor meget memory den bruger når den kører.

Vis med skærmdump hvor mongo-serverens logfil ligger.

Opgave 2 – Beskrivende statistik og visualisering

Opgave 2.1 – Afleveringer

Find antal afleveringer vha en mongo-query. I skulle gerne lande på lidt over 1 million. Hvis jeres maskiner ikke kan trække det i én dataframe kan I med fordel benytte jer af filtre i jeres mongo-queries. Men husk at få alle records fra kvinde-kampene med.

Lav en optælling af antal fejlafterleveringer fordelt på køn. Udregn f.eks. hvor mange procent fejlafterleveringer udgør af de samlede skud for begge køn. Gør det samme for skuddene – altså hvor mange procent resulterer i mål.

Lav samme type af udregning i forhold til længden af afleveringerne.

Opgave 2.2 – Skud

Lave en top ti for henholdsvis mænd og kvinder på antal skud.

Opgave 3 – Kvinder, mænd og etik

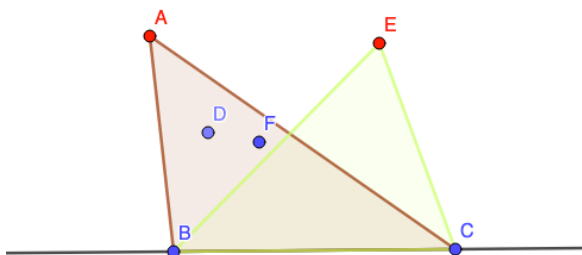
Opgave 3.1 – Forskel i events mellem kvinder og mænd

Er der forskel på kvinder og mænd, når det gælder events i fodboldkampe?

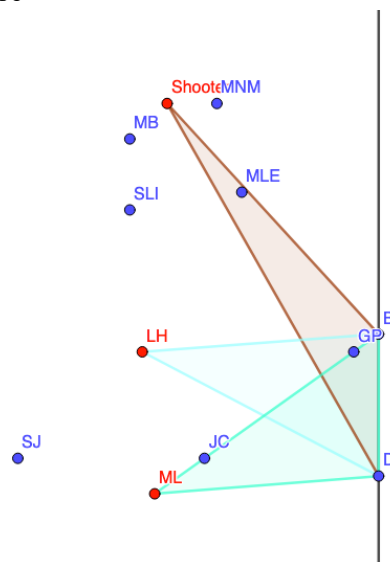
Overvej om der er måder man kan analysere data som vil give et mere nuanceret billede af en kamp. Det kunne f.eks. være i forbindelse med skud, hvor man kiggede på freeze-ramen for at se hvor ofte en medspiller er i en bedre position i forhold til målet end den, der skyder. Antagelsen skulle være at kvinder har et stærkere blik for helheden og vil derfor være bedre til at aflevere bolden såfremt en medspiller har bedre chancer for at score.

Man kunne starte med at tælle hvor mange modspillere der befinder sig indenfor den trekant, der dannes mellem spilleren og målets to stolper. Så f.eks. er spiller E i bedre position end spiller A. Matematikken kan I hente fra f.eks. denne video: <https://www.youtube.com/watch?v=HYAgJN3x4GA> eller google "mathematics point in triangle."

Bedre position



Test Case



```
Test: db.matches.find({"_id":"b20d5df4-6c96-4801-bd59-2ac6ca90aa5b"})
```

Georgia Stanway vælger at skyde og skuddet reddes af GP. Ud fra det simple mål (antal modspillere i din triangle) burde hun have afleveret til LH som kun har 1 modspiller i sin triangle.

Min python-kode giver flg på denne test-række i data-ramen med alle skud:

```
'ROW b20d5df4-6c96-4801-bd59-2ac6ca90aa5b'  
Georgia Stanway has 2 in tri  
{'id': 15547, 'name': 'Melissa Lawley'} has 2 in tri  
{'id': 15555, 'name': 'Lauren Hemp'} has 1 in tri  
CALLED ...
```

Opgave 3.2 – Investering i kvindefodbold

Kvindefodbold er i disse år i markant fremgang. Argumentér ud fra jeres data om I ville investere i kvindefodbold.

Opgave 3.3 – Etiske overvejelser

Overvej hvordan man kunne argumentere for at indføre blandede hold. Hvordan kunne man rent praktisk gøre det (altså hvordan skulle man ændre reglerne).

Opstil en lille matrix og udfyld felterne med de overvejelser man ville gøre sig i de tre lejre – givet at man tog en principiel beslutning om at indførte blandede ligaer.

	Social constructivism	Essentialism	Egalitarianism
Mix male and female players			
Equal pay for male and female players			